WA1 – 10:45

ROBUSTNESS OF MULTILOOP LINEAR FEEDBACK SYSTEMS*

J.C. Doyle
Consultant
Honeywell Systems and Research Center
2600 Ridgway Parkway
Minneapolis, Minnesota 55413

ABSTRACT

This paper presents a new approach to the frequency-domain analysis of multiloop linear feedback systems. The properties of the return difference equation are examined using the concepts of singular values, singular vectors and the spectral norm of a matrix. A number of new tools for multiloop systems are developed which are analogous to those for scalar Nyquist and Bode analysis. These provide a generalization of the scalar frequency-domain notions such as gain, bandwidth, stability margins and M-circles, and provide considerable insight into system robustness.

1. Introduction

A critical property of feedback systems is their robustness; that is, their ability to maintain performance in the face of uncertainties. In particular, it is important that a closed-loop system remain stable despite differences between the model used for design and the actual plant. These differences result from variations in modelled parameters as well as plant elements which are either approximated, aggregated, or ignored in the design model. The robustness requirements of a linear feedback design are often specified in terms of desired gain and phase margins and bandwidth limitations associated with loops broken at the input to the plant actuators ([1], [2]). These specifications reflect in part the classical notion of designing controllers which are adequate for a set of plants constituting a frequency-domain envelope of transfer functions [3]. The bandwidth limitation provides insurance against the uncertainty which grows with frequency due to unmodeled or aggregated high frequency dynamics.

Thus, the classical notions of gain, bandwidth, and stability margins have played a critical role in evaluating system robustness. The Bode plot or scalar Nyquist or Inverse Nyquist diagram (polar plots of the loop transfer function) provides a means of assessing these quantities at a glance. For multiloop systems, individual loops may be analyzed using scalar techniques, but these methods fail to capture the essentially multivariable nature of many systems. For example, scalar methods may ignore variations which simultaneously affect multiple loops.

There are a number of other possible ways to extend the classical frequency-domain techniques. One involves using compensation or feedback to decouple (or approximately decouple) a multiloop system into a set of scalar systems which may be treated with scalar techniques (i.e., "Diagonal Dominance", Rosenbrock [4]). Another method uses the eigenvalues of the loop transfer matrix (G(s) in Figure 1) as a function of frequency (i.e., "Characteristic Loci", MacFarlane, et. al. [5], [6]). While these methods may provide legitimate tools for dealing with some multivariable systems, they can lead to highly optimistic conclusions about the robustness of multiloop feedback designs. Examples in Section III will demonstrate this.

This paper develops an alternative view of multiloop feedback systems which exploits the concepts of singular values, singular vectors, and the spectral norm of a matrix. ([7] - [10]). This approach leads to a reliable method for analyzing the robustness of multivariable systems.

Section 2 presents a basic theorem on robustness and sensitivity properties of linear multiloop feedback systems. Multivariable generalizations of the scalar Nyquist, Inverse Nyquist and Bode analysis methods are then developed from this same result.

Two simple examples are analyzed in Section 3 using the tools of Section 2. As promised, the inadequacies of the existing approaches outlined earlier will be made clear.

Section 4 contains a discussion of some of the implications of this work.

The goal of this paper is to focus on the analysis of robustness and sensitivity aspects of linear multiloop feedback systems. Some new

* This work was performed for Honeywell Systems and Research Center, Minneapolis, MN, with partial support from the Office of Naval Research Contract N-00014-75-C-0144

Mailing Address: MS MN17-2367, Honeywell Systems and Research Center
2600 Ridgway Parkway, Minneapolis, MN 55413

approaches emerge which yield important insights into their behavior. The mathematical aspects of these topics are fairly mundane at best, so rigor and generality are most always sacrificed for simplicity.

Preliminaries and Definitions

A brief discussion of singular values and vectors follows. Although the concepts apply more generally, only square matrices will be considered in this paper. A more thorough discussion of these topics may be found in [7] - [10].

The singular values σ_i of a complex n x n matrix A are the non-negative square roots of the eigenvalues of A^*A where A^* is the conjugate transpose of A. Since A^*A is Hermitian, its eigenvalues are real. The (right) eigenvectors v_i of A^*A and u_i of AA^* are the right and left singular vectors, respectively, of A. These may be chosen such that

$$\begin{aligned} &\sigma_i u_i = Av_i, \qquad i = 1,\ldots n \\ &\sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_n \end{aligned} \tag{1}$$

and the $\{u_i\}$ and $\{v_i\}$ form orthonormal sets of vectors.

It is well known that

$$A = U \Sigma V^* \tag{2}$$

Where U and V consist of the left and right singular vectors, respectively, and Σ = diag. $(\sigma_1,\ldots,\sigma_n)$. The decomposition in (2) is called the singular value decomposition.

Denote

$$\underline{\sigma}(A) = \min_{||x||=1} ||Ax|| = \sigma_1 \tag{3}$$

and

$$\overline{\sigma}(A) = \max_{||x||=1} ||Ax|| = ||A||_2 = \sigma_n \tag{4}$$

The singular values are important in that they characterize the effect that A has as a mapping on the magnitude of the vectors **x**. They may be thought of as generalizing to matrices the notion of gain. The singular values also give a measure of how "close A is to being singular (in a parametric sense). In fact, the quantity

$$\underline{\sigma} / \overline{\sigma}$$

is known as the condition number with respect to inversion [9]. The eigenvalues of A do not in general give such information. If λ is an eigenvalue of A, then

$$\underline{\sigma} \leq |\lambda| \leq \overline{\sigma}$$

and it is possible for the smallest eigenvalue to be much larger than $\underline{\sigma}$.

2. Basic Results

Consider the feedback system in Fig. 2 where G(s) is the rational loop transfer matrix and L(s) is a perturbation matrix, nominally zero, which represents the deviation of G(s) from the true plant. While this deviation is unknown, there is usually some knowledge as to its size. A reasonable measure of robustness for a feedback system is the magnitude of the otherwise arbitrary perturbation which may be tolerated without instability. The following theorem characterizes robustness in this way. The "magnitude" of L(s) is taken to be the spectral norm. Only stable perturbations are considered since no feedback design may be made robust with respect to arbitrary unmodeled unstable poles.*

Robustness theorem: Consider the perturbed system in Figure 2 with the following assumptions

i) G(s) and L(s) are nxn rational square matrices

ii) det $(G(s)) \not\equiv 0$

iii) L(s) is stable

iv) the nominal closed loop system

$H = G(I+G)^{-1}$
is stable.

Under these assumptions the perturbed system is stable if

$$\underline{\sigma}\,(I+G(s)^{-1}) > \overline{\sigma}\,(L(s)) \tag{5}$$

for all s in the classical Nyquist D-contour (defined below)

Proof:

It is well known [4] that since G is invertible

$$\det(H(s)^{-1}) = \det\,(I+G(s)^{-1}) = \frac{\psi_1(s)}{\phi_1(s)} \tag{6}$$

where $\psi_1(s)$ is the nominal closed-loop characteristic polynomial and $\phi_1(s)$ is the transmission zero polynomial of G [11].

For the perturbed system

$$\det\,(I + G(s)^{-1} + L(s)) = \frac{\psi_2(s)}{\phi_1(s)\psi_3(s)} \tag{7}$$

where $\psi_2(s)$ is the perturbed closed-loop characteristic polynomial and $\psi_3(s)$ is the characteristic polynomial of L(s).

Let D be a large contour in the s-plane consisting of the imaginary axis from -jR to +jR, together with a semicircle of radious R in the right half-plane. The radious R is chosen large enough so that all finite roots of $\psi_2(s)$ have magnitude less than R.

*It is possible that this requirement may be relaxed somewhat. See Section 4.

Let the contour Γ_o be the image of D under the map $\psi_1(s)$ det $(I + {}^o G(s)^{-1})$. Since H is stable, it follows from the principle of the argument that Γ_o will not encircle the origin.

Define the map

$$\gamma(q,s)=\phi_1(s) \det (I+G(s)^{-1}+qL(s)), \text{ q real} \tag{8}$$

and let $\gamma(q,s)$ map D into the Contour $\Gamma(q)$ for fixed q, $o\leq q\leq 1$. The map $\gamma(q,s)$ may be written as

$$\gamma(q,s)= \frac{\psi_1(s)\psi_3(s) + q\theta_1(s)+\ldots+q^n\theta_n(s)}{\psi_3(s)}$$

$$= \frac{\psi_4(q,s)}{\psi_3(s)} \tag{9}$$

Clearly, since $\Gamma(o) = \Gamma_o$, it does not encircle the origin. Since the roots of ψ_4 are algebraic functions of q, they are continuous in q [12]. Thus the only way that the perturbed contour $\Gamma(1)$ can encircle the origin is for

$$\det (I+G(s)^{-1} + qL(s)) = 0 \tag{10}$$

for some s in D and some q on the interval o<q<1. (Recall that $\psi_3(s)$ has no right half-plane roots). When (10) is satisfied then $\underline{\sigma}$ $(I+G^{-1} + qL)$ must also be zero. However, as a consequence of (5)

$$\underline{\sigma}\ (I+G^{-1}+qL) \geq \underline{\sigma}\ (I+G^{-1}) - q\ \overline{\sigma}(L)$$

$$\geq \underline{\sigma}\ (I+G^{-1}) - \overline{\sigma}\ (L)$$

$$> 0 \tag{11}$$

Thus $\Gamma(q)$ does not encircle the origin for $o\leq q\leq 1$. In particular, the perturbed contour $\Gamma(1)$ does not encircle the origin, and the perturbed closed-loop system is stable. □

Similar theorems hold for additive rather than multiplicative perturbations (with I + G substituted for $I + G^{-1}$) as well as a number of other configurations.

This theorem points out the importance of singular values. In particular, the smallest singular value $\underline{\sigma}(I + G(j\omega)^{-1})$ gives a reliable frequency-dependent measure of robustness. Stability is guaranteed for all perturbations whose spectral norm is less than $\underline{\sigma}$. As will be seen in the examples, eigenvalues do not give a similarly reliable measure.

The singular values also have useful graphical interpretations. Consider the dyadic expansion

$$H^{-1} = I + G^{-1} = \sum_{i=1}^{n} \sigma_i u_i v_i^* \tag{12}$$

$$\sigma_1\leq_2\leq\cdots\leq\sigma_n$$

where the σ_i, u_i and v_i are the singular values, and left and right singular vectors, respectively of $I + G^{-1}$. This is an alternative form of the singular value decomposition in equation (2).

The quantities σ_i, u_i and v_i may be viewed as functions of a complex variable, or in particular as functions of frequency.

Since

$$H = (I + G^{-1})^{-1} = \Sigma \frac{1}{\sigma_i} v_i u_i^* \tag{13}$$

the values $1/\sigma_1(j\omega)$ and $1/\sigma_n(j\omega)$ give the maximum and minimum possible magnitude responses to an input sinusoid at frequency ω. Eigenvalues give no such information. In this sense, a plot of these singular values vs. frequency may be thought of as a multivariable generalization of the Bode gain plot. Plots of this type will be referred to as σ-plots. Another useful graphical interpretation analogous to the scalar Inverse Nyquist diagram may be constructed by noting that

$$G^{-1} = \Sigma\ \sigma_i \ddot{u}_i v_i^* - I$$

$$= \Sigma\ \beta_i g_i v_i^* \tag{14}$$

where $\beta_i g_i = \sigma_i u_i - v_i$ with β_i real and $||g_i|| = 1$ for all i. (The g_i's do not necessarily form an orthonormal set.)

The quantities in (14) at some frequency ω_o are related as in diagram in Fig. 3a. Since v_i is of unit length a complex plane may be constructed as in Fig. 3b, to lie in the plane formed by the triangle of v_i, $\sigma_i u_i$ and $\beta_i g_i$.

Define z_i to be the complex number at the point of the triangle as in Fig. 3c. Then, by rotating the complex plane with the triangle as a function of frequency, a $z_i(j\omega)$ may be obtained which is a continuous function of ω (Fig. 3d). This allows the important quantities in (13) and (14), that is, the σ_i and β_i to be represented in convenient graphical form. As noted in Fig. 3d, there is an ambiguity to z_i depending on which side the plane is viewed. (To be more precise, the z_i represent a multivalued function of s which could be defined on appropriate Riemann sheets. However, this will be ignored.) The z_i may be calculated by finding the roots of the quadratic equation

$$z_i^{\,2} + (1+\beta_i^{\,2}-\sigma_i^{\,2})z_i + \beta_i^{\,2} = 0 \tag{15}$$

By plotting the $z_i(j\omega)$ $(i = 1,\ldots m)$ for frequencies of interest a plot analogous to the scalar Inverse Nyquist plot is generated. While phase does not have the conventional meaning on these plots, the more important notion of distance from the critical point preserves its importance. Gain and bandwidth may also be interpreted in the conventional manner. These plots will be referred to as z-plots.

Concepts such as M-circles are also obvious in this context. The minimum value of M is given by

$$M_m = \max_{\omega} (1/\sigma_1(j\omega))$$

Similar results may be obtained for additive perturbations by working with $I + G$ rather than $I + G^{-1}$. In this case a diagram is generated which is analogous to the scalar Nyquist diagram. A number of other configurations may be handled as well.

Note that singular values offer no encirclement condition to test for right half-plane poles. Another test must be made for absolute stability but this presents no obstacle as many simple techniques exist for doing this. Once stability is determined the various approaches presented in this Section may be used to reliably analyze robustness.

3. Examples

Two simple examples are presented and analyzed using the approaches developed in the previous section. For the purpose of comparison, the methods of loop-breaking, direct eigenvalue analysis of G, and diagonalization by compensation are also used. The advantage of the interpretation of robustness given in this paper is clearly illustrated.

The first example is an oscillator with open loop poles at $\pm 10j$ and both closed loop poles at -1. There are no transmission zeros. The loop transfer function is

$$G(s) = \frac{1}{s^2+100} \begin{bmatrix} s-100 & 10(s+1) \\ -10(s+1) & s-100 \end{bmatrix} \tag{16}$$

By closing either loop (the system is symmetric) as in Figure 4, the transfer function for the other loop is

$$g(s) = \frac{1}{s}$$

which indicates ∞ db gain margin in both directions and 90° phase margin in each loop (with the other closed). This is very misleading, however.

The z-plot for this example is shown in Figure 5. It may appear somewhat peculiar, since it is not a plot of a rational function. The important feature is the proximity of the plot to the critical point, indicating a lack of robustness.

The apparent discrepency between these two robustness indications can be easily understood by considering a diagonal perturbation

$$L = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix} \tag{17}$$

where k_1 and k_2 are constants.

Then regions of stability and instability may be plotted in the (k_1, k_2) plane as has been done in Figure 6. The open loop point corresponds to $k_1 = k_2 = -1$ and nominal closed loop point point corresponds to $k_1 = k_2 = 0$. Breaking each loop individually examines stability along the k_1, k_2 axes where robustness is good, but misses the close unstable regions caused by simultaneous changes in k_1 and k_2. Thus, single loop analysis is not a reliable way of testing robustness.

The second example is a two dimensional feedback system with open loop poles at -1 and -2 and no transmission zeroes.

The loop transfer matrix is

$$G(s) = \frac{1}{(s+1)(s+2)} \begin{bmatrix} -47s + 2 & 56s \\ -42s & 50s + 2 \end{bmatrix} \tag{18}$$

Assume that identity feedback is used, with closed-loop poles at -2 and -4. This system may be diagonalized by introducing constant compensation. Let

$$U = \begin{bmatrix} 7 & 8 \\ 6 & 7 \end{bmatrix} \tag{19}$$

and

$$V = U^{-1} = \begin{bmatrix} 7 & -8 \\ -6 & 7 \end{bmatrix} \tag{20}$$

Then letting

$$\hat{G} = VGU = \begin{bmatrix} \frac{1}{s+1} & 0 \\ 0 & \frac{2}{s+2} \end{bmatrix} \tag{21}$$

the system may be rearranged so that

$$\begin{aligned} H &= G(I+G)^{-1} \\ &= U\hat{G}V(I + U\hat{G}V)^{-1} \\ &= U\hat{G}(I + \hat{G})^{-1}V \\ &= U[\hat{G}(I + \hat{G})^{-1}]V. \end{aligned}$$

This yields a diagonal system that may be analyzed by scalar methods. In particular under the assumption of identity feedback $\hat{G}$ represents the new loop transfer matrix. Because U and V represent a similarity transformation, the diagonal elements of G are also the eigenvalues of G so that the decoupling or dominance approach and eigenvalue or characteristic loci approach would generate the same Nyquist or Inverse Nyquist plot shown in Figure 7. Only a single locus is shown since the contours of $1/(s+1)$ and $2/(s+2)$ are identical. The tempting conclusion that might be reached from these plots is that the feedback system is emminently robust with apparent margins of $\pm\infty$ db in gain and 90° in phase. The closed-loop pole locations would seem to support this.

This conclusion, however, would be wrong. The z-plot for $I + G^{-1}$ is shown in Figure 8 and there is clearly a serious lack of robustness. The (k_1, k_2) - plane stability plot for this example is shown in Figure 9. Neither the diagonal dominance nor eigenvalue approaches indicate the close proximity of an unstable region. This failure can be attributed to two causes.

First, the eigenvalues of a matrix do not, in general, give a reliable measure of its distance (in a parametric sense) from singularity, and so computing the eigenvalues of $G(s)$(or $I+G(s)$) does not give an indication of robustness. Using eigenvalues rather than singular values will always detect unstable regions that lie along the $k_1 = k_2$ diagonal, but may miss regions such as the one in Figure 9.

Second, when compensation and/or feedback is used to achieve dominance, the "new plant" includes this compensation and feedback. Because of this, no reliable conclusions may be drawn from this "new plant" concerning the robustness of the final design with respect to variations in the actual plant. It is important to evaluate robustness where there is uncertainty.

Another important property of multiloop feedback is that, unlike scalar feedback, pole locations alone are not reliable indicators of robustness. This was demonstrated in the last example and may be explained as follows. Consider a state feedback problem where the plant is controllable from each of two inputs. One input may be used to place the poles far into the right half plane and the other used to bring them back to the desired location. Such a high-gain control design of "opposing" loops will be extremely sensitive to parameter variations regardless of the nominal pole locations.

It is interesting to examine the σ-plot of $H = G(I + G)^{-1}$ for the second example shown in Figure 10. Recall that the singular values of H are equal to the inverses of the singular values of $I + G^{-1}$. There is a rather large peak in the frequency response at approximately 3 radians. This could not occur in scalar unity feedback systems without there being a pole relatively near the imaginary axis. It can happen in multiloop systems because of the high gains possible without corresponding large pole movement.

4. Further Comments and Conclusions

The approach to the analysis of robustness presented here appears to yield useful insight into the properties of multiloop feedback systems. One possible difficulty with the approach is that it can lead to overly pessimistic views of robustness because it considers perturbations which may not be physically possible. This problem exists as well with gain and phase margin evaluations. Of course, some of this difficulty can be alleviated by examining the specific perturbations leading to instability. These may be easily computed from equation (12). On the other hand, it might be argued that a little healthy pessimism would be refreshing in the field of multivariable linear control research.

The results in this paper concerning dominance methods and use of characteristic loci of the loop transfer matrix are not meant to imply that design procedures employing these methods are useless. However, simply designing "in the frequency domain" is no guarantee that resulting controllers will have no undesirable properties.

Although for simplicity's sake only rational transfer functions were considered the results in this paper should extend to nonrational transfer functions. In practical application it should be possible to use frequency response data directly. The results may also be extended to include nonlinear perturbations by exploiting the general stability theory developed by Safonov [13]*. In this setting, nonlinearities may be loosely viewed as linear time-invariant elements with time-varying parameters. A mathematically more rigorous treatment of these issues may be found in Zames ([14], [15]), as well as in [13].

A limitation to the robustness theorem as stated is the requirement that L(s) be stable. In practice, it is not uncommon for physical systems to have parameter variations which cause poles to migrate across the imaginary axis. It appears likely that if such pole movements are restricted to some region of the complex plane, and another restriction is made on the system zeroes, a modified robustness theorem may be obtained. This would depend, of course, on the nature of the other sources of system uncertainty as well as the nature of the feedback employed.

The significance of the approach presented herein seems to be in the natural way it gives multivariable interpretations to many important classical control concepts. Preliminary multivariable feedback designs using methods based on these results are most encouraging [18], and this too appears to be a promising area for research.

*Recently, it has been shown that the results in this paper may be derived from those in [13], and the robustness results in [13] can be expressed in terms of singular values [17].

Acknowledgements

I would like to thank all those whose criticisms and comments helped to mold this paper. I would particularly like to thank Drs. G. Stein, M.G. Safonov, and C.A. Harvey for their continued technical input.

I would also like to thank the Math Lab Group, Laboratory for Computer Sciences, MIT for use of their invaluable tool, MACSYMA, a large symbolic manipulation language. The Math Lab Group is supported by NASA under grant NSG 1323 and DOE under contract #E(11-1)-3070.

References

[1] B.C. Kuo, Automatic Control Systems, Prentice-Hall, 1967.

[2] J.W. Brewer, Control Systems, Prentice-Hall, 1974.

[3] I.M. Horowitz, Synthesis of Feedback Systems, Academic Press, 1963.

[4] H.H. Rosenbrock, Computer-Aided Control System Design, Academic Press, 1974.

[5] A.G.J. MacFarlane and I. Postlethwaite, "The Generalized Nyquist Stability Criterion and Multivariable Root Loci", Int. J. Control, Vol. 23, No. 1, January, 1977, pp. 81-128.

[6] A.G.J. MacFarlane and B. Kouvaritakis, "A Design Technique for Linear Multivariable Feedback Systems", Int. J. Control, Vol. 23, No. 6, June, 1977, pp. 837-874.

[7] G. Strang, Linear Algebra and It's Applications, Academic Press, 1976.

[8] J.H. Wilkinson and C. Reinsch, Linear Algebra, Springer-Verlag, 1971.

[9] J.H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, 1965.

[10] G.E. Forsythe and C.B. Moler, Computer Solutions of Linear Algebraic Systems, Prentice-Hall, 1967.

[11] A.G.J. MacFarlane and M. Karcanias, "Poles and Zeros of Linear Multivariable Systems: A Survey of Algebraic, Geometric, and Complex Variable Theory", Int. J. Control, July, 1976, Vol. 24, No. 1, pp. 33-74.

[12] K. Knopp, Theory of Functions, Dover, 1947.

[13] M.G. Safonov, "Robustness and Stability Aspects of Stochastic Multivariable Feedback System Design", Ph.D. dissertation, Mass. Inst. Tech., Sept. 1977.

[14] G. Zames, "On the Input-Output Stability of Time-Varying Nonlinear Feedback Systems - Part I", IEEE Tran. on Automatic Control, Vol. AC-11, No. 2, pp.228-238, Apr. 1966.

[15] G. Zames, "On the Input-Output Stability of Time-Varying Nonlinear Feedback Systems - Part II", IEEE Trans. On Automatic Control, Vol. AC-11, No. 3, pp. 465-476, July, 1966.

[16] B.S. Garbo, et al, Matrix Eigensystems Routine - EISPACK Guide Extension, Lecture Notes in Computer Science, Volume 51, Springer-Verlag, Berlin, 1977.

[17] M.G. Safonov, "Singular Values and Multi-loop Nonlinear System Robustness", Honeywell Memo MR 12513.

[18] G. Stein and J. Doyle, "Singular Values and Feedback: Design Examples", 16th Annual Allerton Conference on Communication, Control, and Computing, Monticello, Ill. October 4-6, 1978.

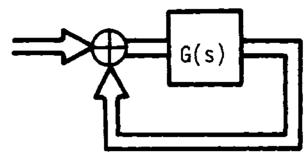

Figure 1. Multiloop Feedback System

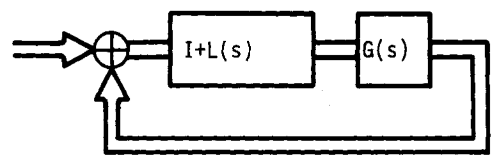

Figure 2. Perturbed System

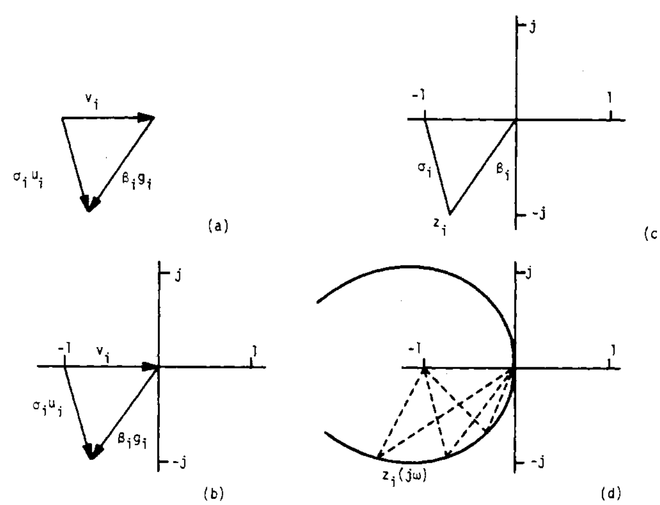

Figure 3 Construction of z-plot

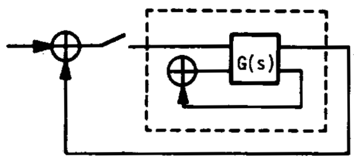

Figure 4. Analysis of Loop-Breaking

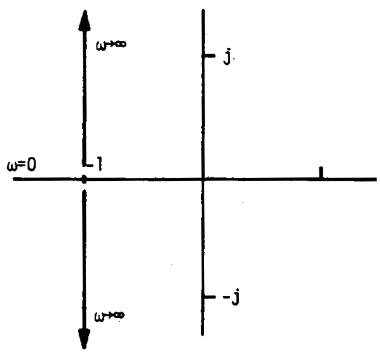

Figure 5. Example 1 z-plot

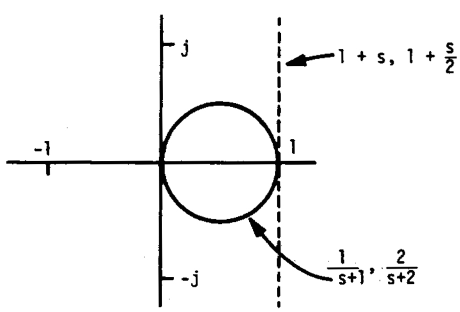

Figure 7. Example Nyquist and Inverse Nyquist Diagram

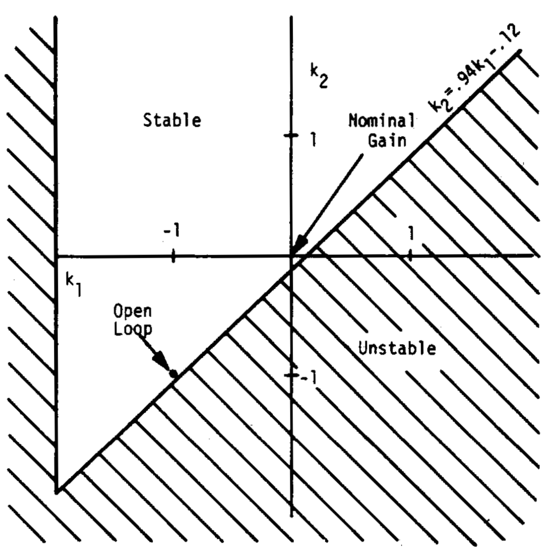

Figure 9. Example 2 Stability Domain

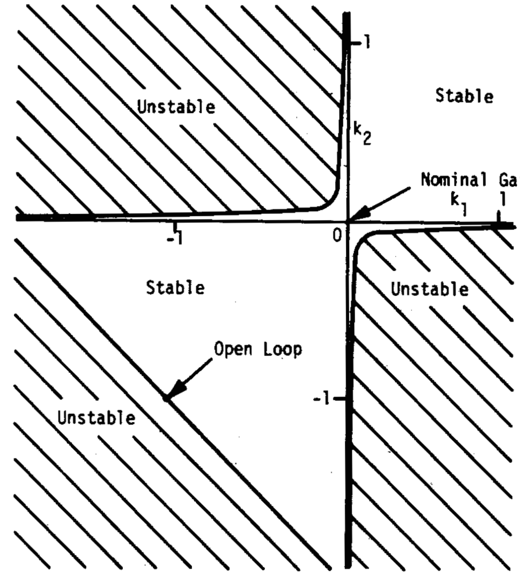

Figure 6. Example 1 Stability Domain

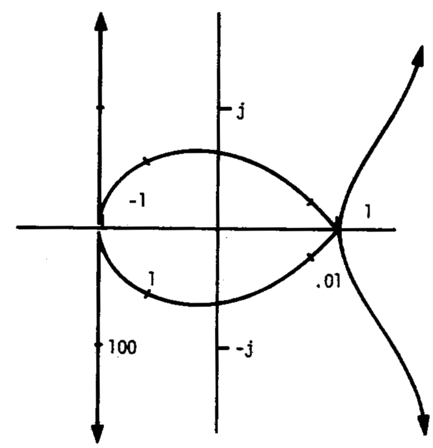

Figure 8. Example 2 z-plot

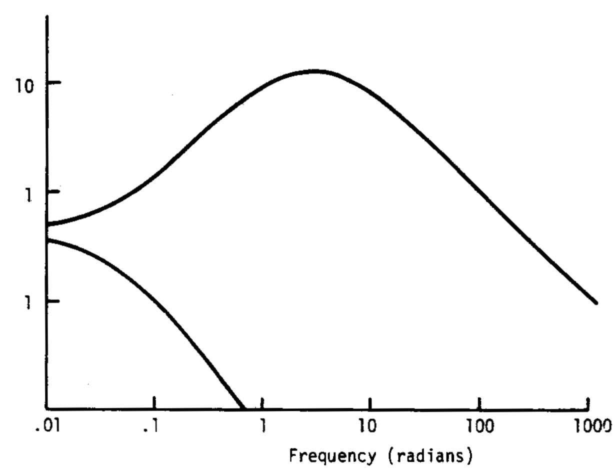

Figure 10. Closed-Loop Frequency Response $-1/\sigma(I+G^{-1})$